

УДК 004.822

Цубера М. – ст. гр. СНм – 51

Тернопільський національний технічний університет імені Івана Пулюя

ВИКОРИСТАННЯ ІНТЕЛЕКТУАЛЬНОГО ПОШУКУ НА САЙТІ

Науковий керівник: асистент Маєвський О. В.

Зараз у зв'язку із тенденцією інтеграції локальних мереж з глобальними, істотно зросла важливість проблеми релевантного пошуку в джерелах різного виду. Проте, добре відомо, що використовувані в існуючих системах методи не дозволяють досягти високої повноти і точності пошуку. Однією з причин є вузька спеціалізація систем пошуку.

У більшості випадків повнота пошуку збільшується за рахунок безперервного моніторингу мережі або бази даних з метою виявлення нових документів. Інший шлях — створення метапошукових систем, що дозволяють об'єднувати результати пошуку з різних джерел.

Для підвищення повноти пошуку використовується метапошукова система, яка дозволяє налаштовуватися на інтерфейси пошукових ресурсів і відправляти перетворені запити користувача відразу на декілька пошукових машин або будь-які інші сайти.

Точність пошуку підвищується за рахунок подальшої обробки інформації і семантичної фільтрації знайдених документів.

Алгоритм роботи системи інтелектуального пошуку:

1. користувач вибирає тип запиту («в мережі Інтернет», «в локальній БД», «в локальній мережі») і вводить пошуковий запит на потрібній мові;

2. запит обробляється, з нього видобуваються ключові слова. При цьому використовується розширення запиту за рахунок використання словника синонімів (використовується словник синонімічних предикатів і словник синонімічних іменних груп), із запиту викидаються стоп-слова і так далі;

3. перетворений таким чином запит відправляється або на декілька пошукових машин Інтернет (наприклад, на Яндекс або Рамблер) або в локальну БД (пошук здійснюється серед поміщених в ній документів), або виконується в папках локальної мережі;

4. знайдені документи обробляються і поміщаються в повнотекстову базу даних системи;

5. запит користувача і знайдені документи піддаються лінгвістичному аналізу, що включає морфологічний, синтаксичний і поверхневий семантичний аналіз, будуються семантичні образи запиту і документів, проводиться порівняння образів і обчислення семантичної релевантності знайдених документів запиту користувача;

6. знайдені документи сортуються відповідно до вичисленої релевантності, низькорелевантні документи відкидаються.

Система складається з декількох компонентів, пов'язаних один з одним. Під компонентом розуміється набір логічно пов'язаних модулів, що мають загальне призначення і є закінченою підсистемою. Основна сполучна ланка компонентів системи — база даних, де централізовано зберігається основна інформація. Для даних, які недоцільно зберігати в реляційній БД, використовуються файлові сховища. Система повинна підтримувати паралельну обробку даних, при цьому використовуватися мультіагентне середовище розподілених обчислень.